group meeting 2022-06-01

Noise reduction for piece-wise linear signals embedded in 2d.
(e.g. biological twitching trajectories)

twitching motility

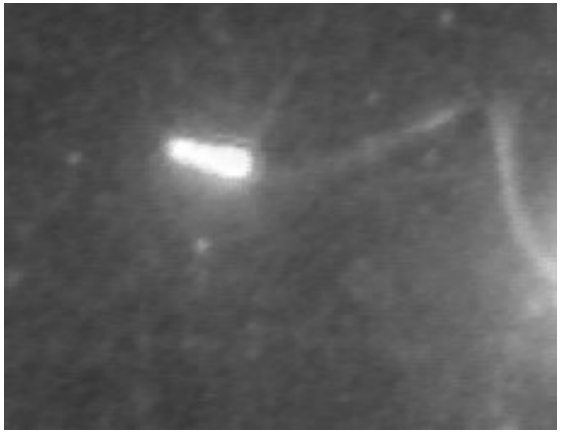


Fig 1. Skerker & Berg, 2001

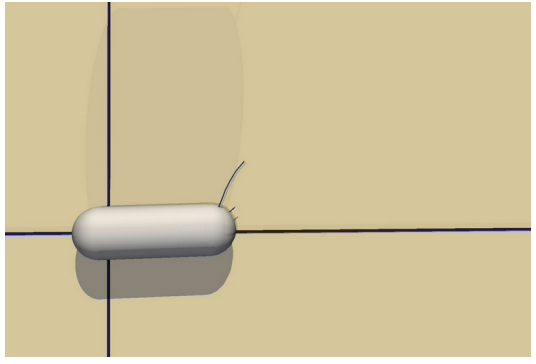


Fig 2. Simulation

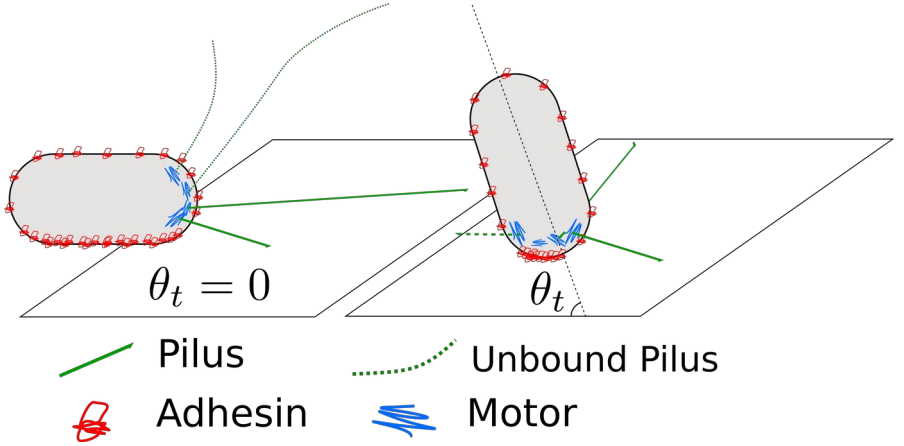


Fig 3. Simplified Drawing

Tracking data from Twitching experiments

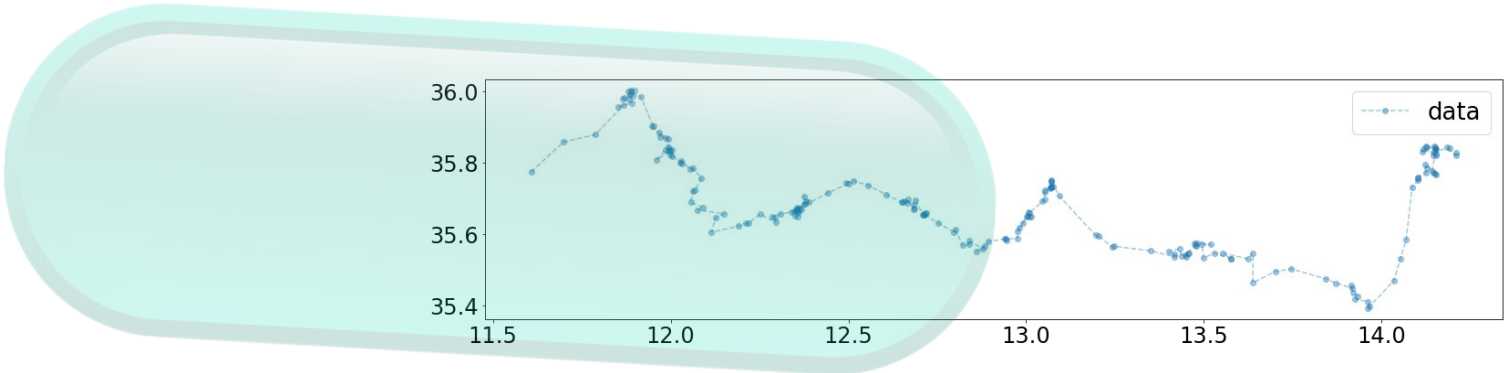


Fig 1. 20 seconds of Tracking Data (μm)

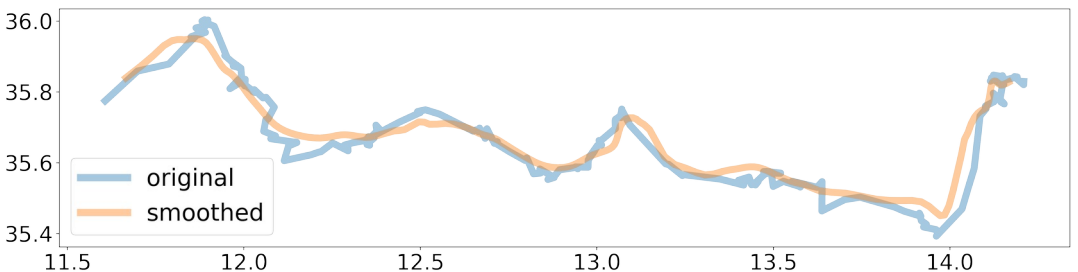
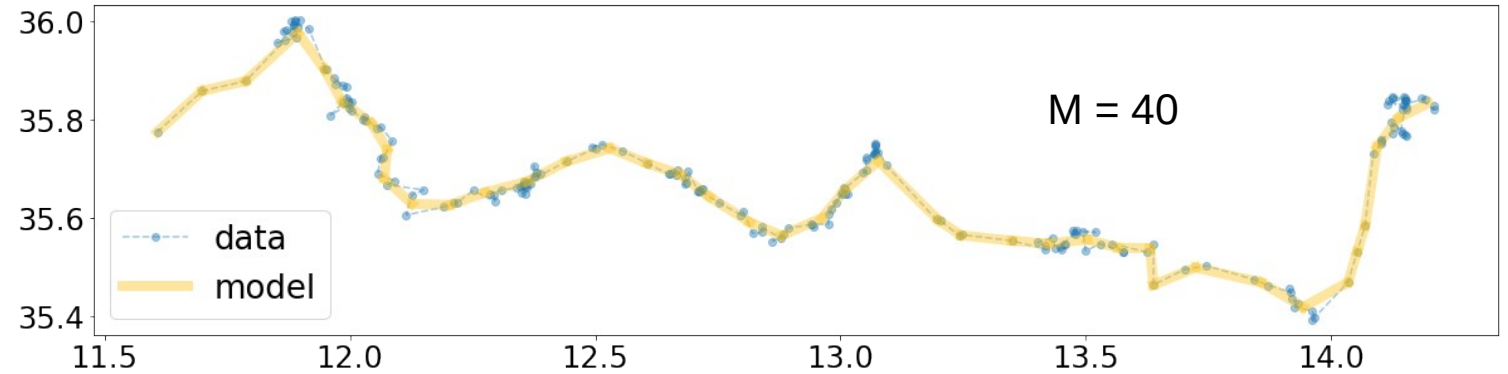
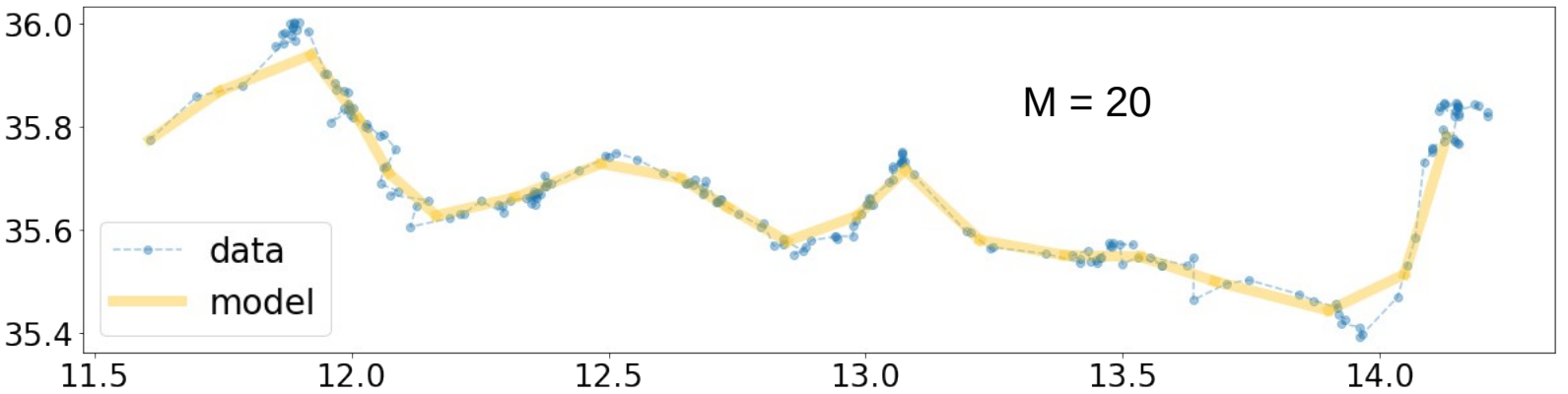


Fig 2. tracking data with wavelet smoothing

Question: Can we get more useful, more interpret-able information from this trajectory data?

What if we only look for piece-wise linear trajectories?





Question: How do we select for the number of segments?

Minimum Description Length Principle

- When choosing an optimal model to describe some data, the best model might be the one which describes the data using the least “information”
  - This approaching is quite natural for humans, e.g. Occam’s Razor
  - Let M be the number of nodes in the piece-wise linear “model”, let N be the number of data points.
  - Each model node is counted a 1 unit of information. M units in total
  - Set a threshold r, points within r of the piece-wise linear curve are counted as part of the model, points further than r are considered outliers and counted separately. Let the number of outliers be n.
  - Description Length (DL) = M + n

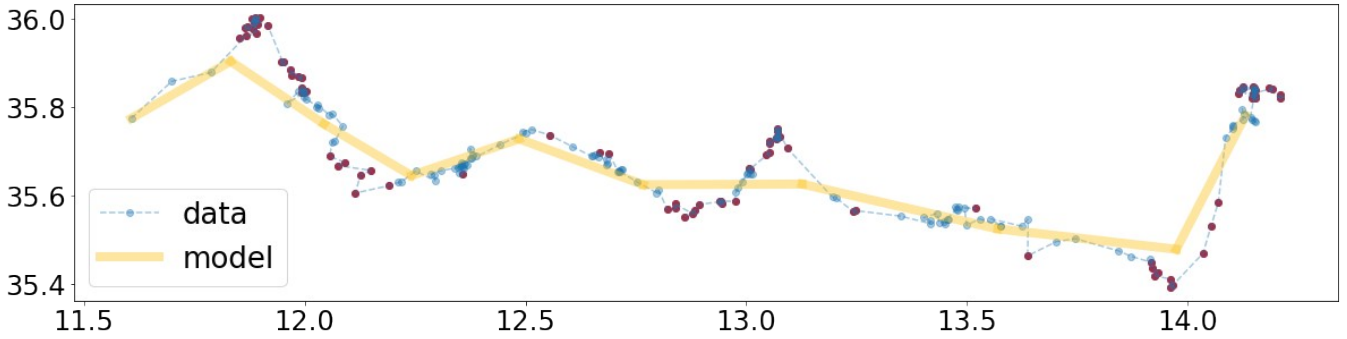
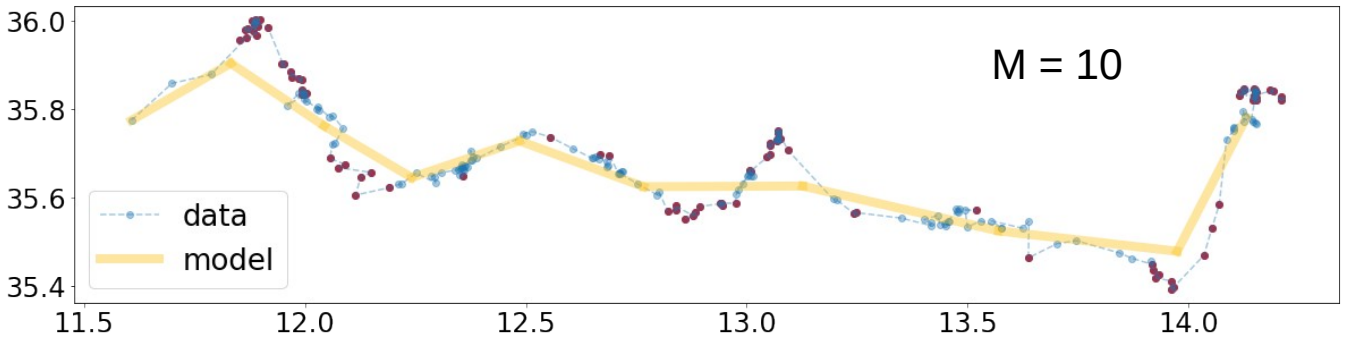


Fig 1. Initial guess with outliers (red) and points close to the model (blue)

Initial guess



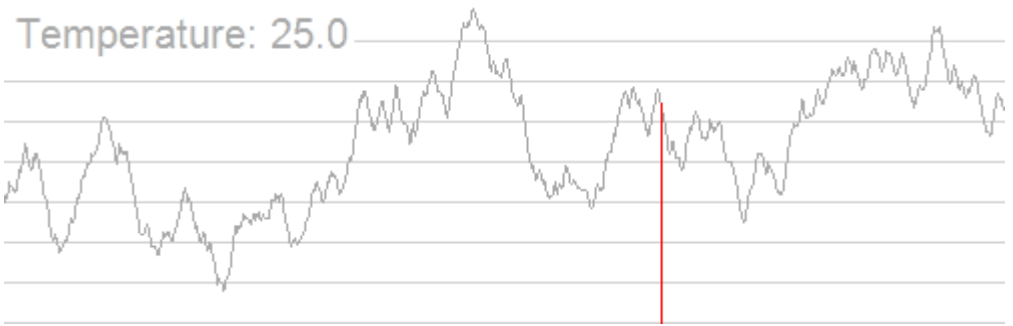
Simple Method for initial guesses.

1. Choose a number of segments M.
2. Identify the pair of adjacent data points with the minimum separation and join them together.
3. repeat until there are M+1 points remaining.

Annealing / Basin Hopping

General Idea

1. randomly choose a conformational change to the piece-wise linear curve, i.e. create or destory a node.
2. Estimate the local minimum of the description length for this conformation
3. If the new description length is better than the current description length, accept the new state of the system, otherwise reject it.
4. go to step 1.



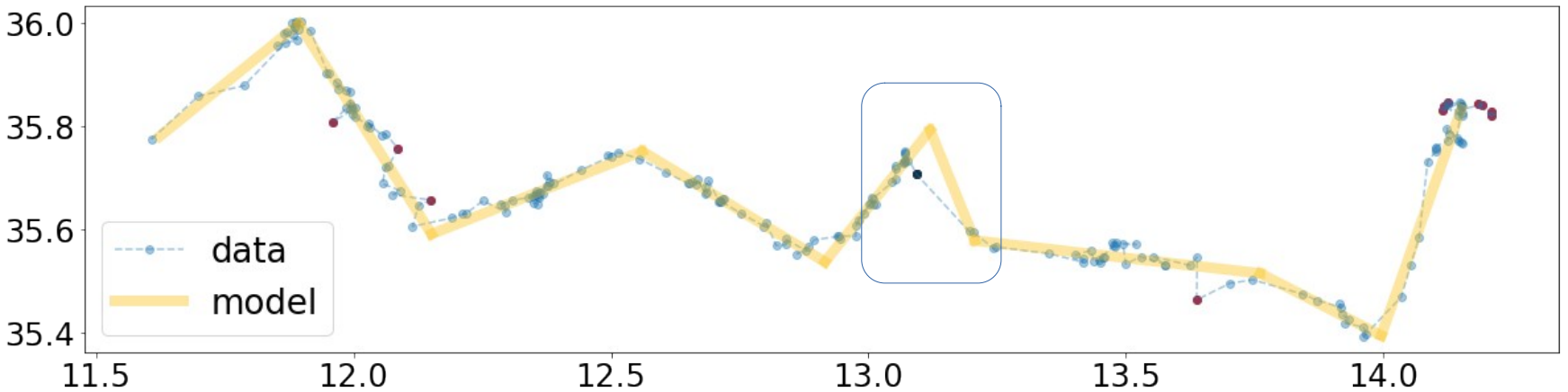
Least Squares

Our least squares function is

$$F = \sum_{i}^{N} \min_{j}(\text{distance}(\boldsymbol{x}_i, \boldsymbol{x}'_j \rightarrow \boldsymbol{x}'_{j+1}))^2$$

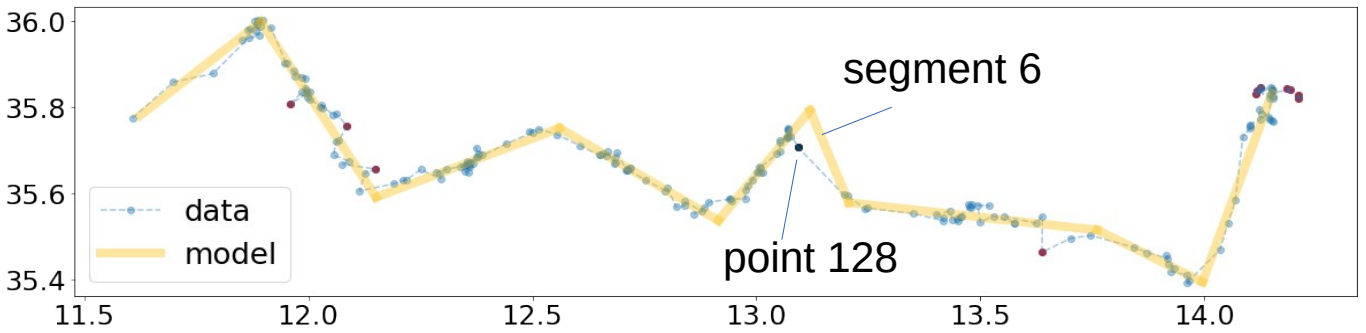
where $\boldsymbol{x}'_j \rightarrow \boldsymbol{x}'_{j+1}$ is the jth PWL segment, let the corresponding vector by denoted by $\boldsymbol{l}_j = \boldsymbol{x}'_{j+1} - \boldsymbol{x}'_j$. We need to evaluate the distance matrix d_{ij},

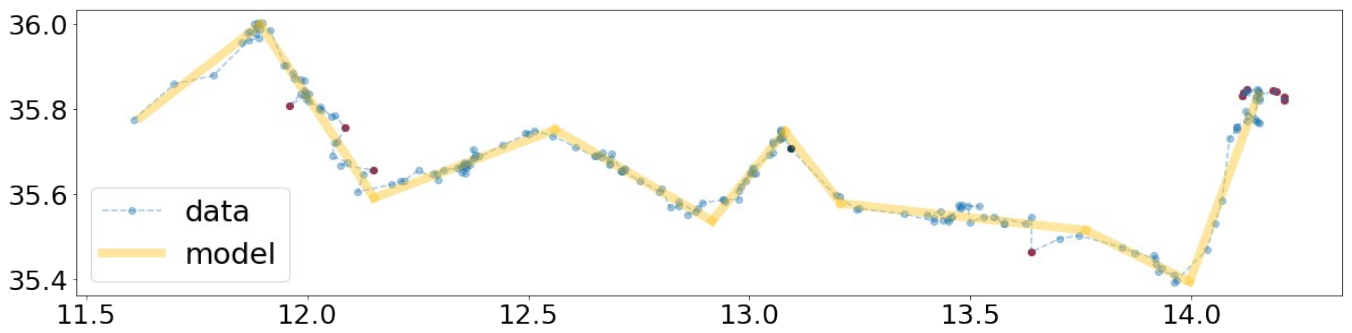
$$d_{ij} = \|\boldsymbol{x}_i - \boldsymbol{x}'_j - s_{ij}\hat{\boldsymbol{l}}_j\|, \quad s_{ij} = \text{clip}((\boldsymbol{x}_i - \boldsymbol{x}'_j) \cdot \hat{\boldsymbol{l}}_j, 0, l_j)$$



An additional Monte Carlo move

- Mapping point 128 to segment 6 leads to a better global minimum





Time ordering term

Let t_i be the time at which the ith data point was collected and s_i be the corresponding curve coordinate we have a series

$$(t_1, s_1), (t_2, s_2), ..., (t_N, s_N)$$

which we sort by s_i to obtain a new series of length N,

$$(t'_1, s'_1), (t'_2, s'_2), ..., (t'_N, s'_N)\,.$$

We then define the term

$$G = \sum_i^N (s'_i - s_i)^2\,,$$

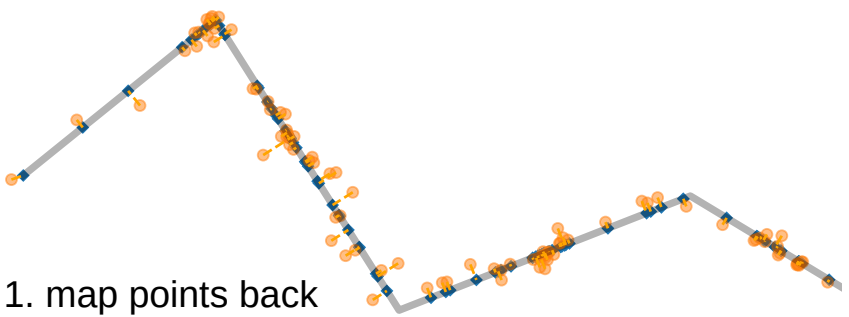


Fig 1. map points back onto curve to get curve coordinates.

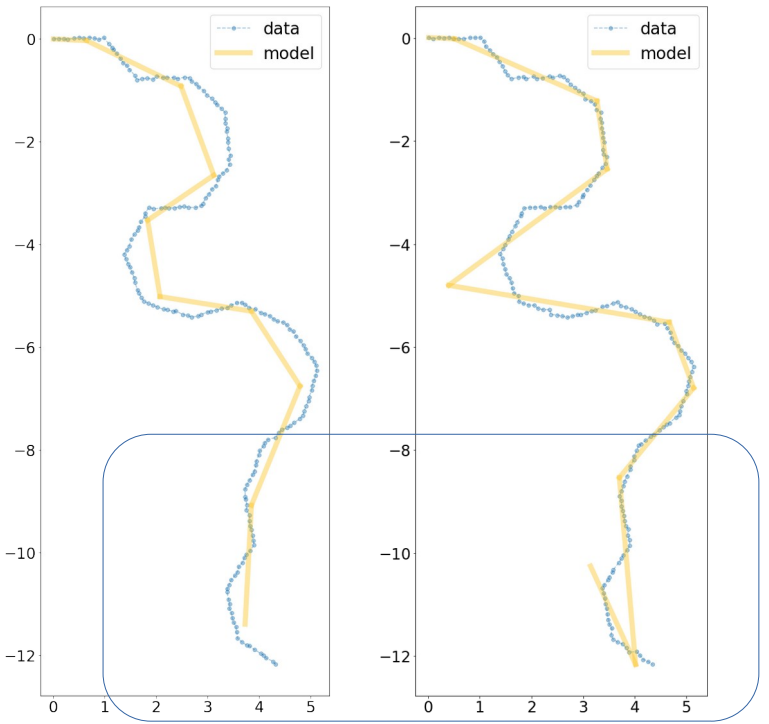
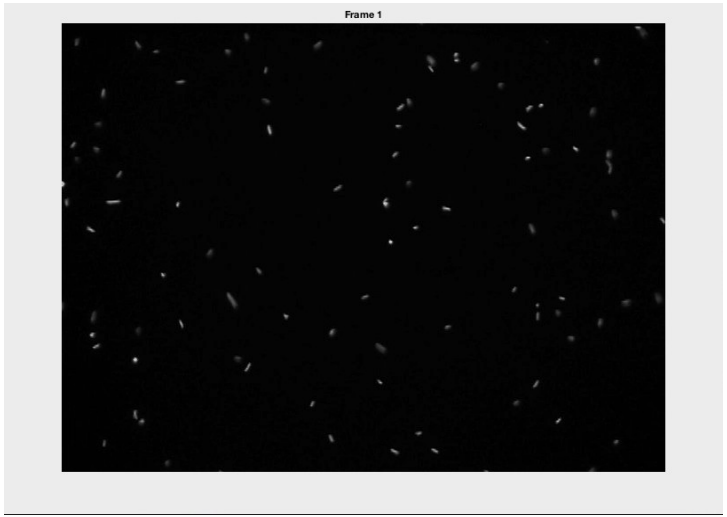


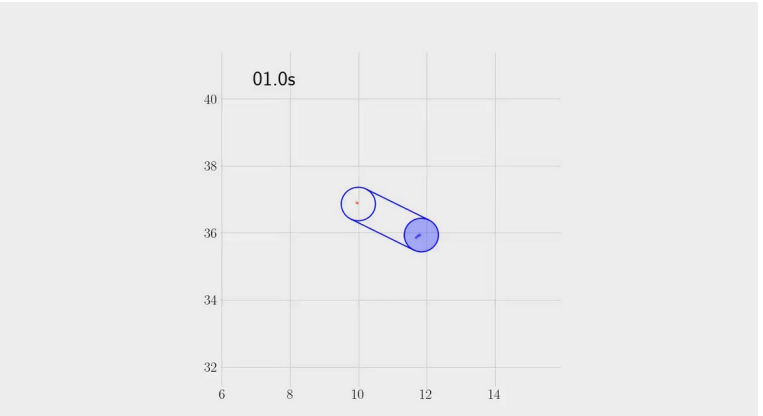
Fig. Initial guess + local solve, with (left) and without (right) additional ordering term.

Todo

- Finish implementation.
- Revise and test implementation using synthetic data.
- Analyse individual trajectories and bacteria populations.



E coli, run and tumble



P. Aeruginosa twitching